Forensic Performance of Short Amplicon Insertion-Deletion (InDel) Markers

Manuel Fondevila Álvarez

Rui Pereira, Leonor Gusmão, Christopher Phillips, John M Butler, María Victoria Lareu, Ángel Carracedo, Peter M Vallone.

22\textsuperscript{nd} International Symposium on Human Identification
Washington D.C. October 5, 2011
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
InDel Polymorphisms

- InDels (insertion-deletion) or DIPs (deletion-insertion polymorphisms) are short length polymorphisms, consisting of the presence or absence of a short (typically 1-50 bp) sequence.

- Closely related to SNPs, sharing most of their properties
  - Low mutation rate – ~$2 \times 10^{-8}$
  - Short amplicon PCR – 60 to 160 bp
  - High multiplexing capacity – 30 to 40 markers

- Total number estimated close to 2 million in the human genome.
As length polymorphisms, InDels can be typed with a simple direct PCR-to-CE genotyping strategy, using a single multiplexed PCR with dyed-linked primers immediately followed by capillary electrophoresis.
Potential Applications of InDels

Degraded DNA samples
- Short amplicon markers

Missing person cases
Mass fatality cases

Complex pedigree kinship
- High multiplexing capacity
- Low mutation rate

Incest cases
Inmigration cases
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
InDel Assays Used in This Study

Qiagen Investigator DIPplex kit

- 31plex PCR
- 30 InDel markers plus amelogenin (on 18 chromosomes)
- Ranging from 75 to 150 bp amplicons

HID-38plex
R. Pereira et al Electrophoresis (2009)

- 38plex PCR
- 38 InDel markers (on 22 chromosomes)
- Ranging from 50 to 155 bp

68 InDel markers in total
9947a DIPplex Profile

Longer InDel fragments leads to interleaving signal

D99 - Deletion
actt____ctctttga

D99 – Insertion of 4 bases
acttTGATctctttga

D97 - Deletion
ggat________tctc

D97 – Insertion of 13 bases
ggatAGAGAAAGCTGAAGtctc

- 30 cycle PCR reaction
- 1 ng of DNA input on PCR
- PCR product diluted 1:250

3000 RFUs

BT matrix standard
9947a HID-38plex profile

Short InDel fragments, No signal interleaving

- **G4 - Deletion**
  
  `actt___ctctttga`

- **G4 – Insertion of 4 bases**
  
  `acttAGGActctttga`

- **R5 - Deletion**
  
  `gga___gtgtca`

- **R5 – Insertion of 4 bases**
  
  `ggaACACgtgtca`

- 29 cycle PCR reaction
- 0.6 ng of DNA input on PCR
- PCR product diluted 1:100
Allele Spread on both InDel Assays

- A shorter Insertion-Deletion fragment length through the assay facilitates the inclusion of a greater number of markers on the same electrophoretic window.
- Moreover, it prevents the interleaving of markers.
Genomic Position of the Markers: Linkage Disequilibrium Possibility

- Indels are thought to play a supporting role to current STR assays.

- Due to the high number of markers, some share the same chromosome region.

- Proximity may pose the risk of markers being linked, if so they could not be statistically multiplied together.

Chromosome 22: A fine example

5 markers are located in close positions in a small region:

- 1 HID-38plex InDel.
- 1 forensic STR.
- 3 DIPplex InDels.

Risk of LD should be evaluated.
When contemplating the possibility of combining the information contained in these InDel markers systems with each other or with core STR loci, we should keep in mind that the proximity between some of these markers could lead to a linkage disequilibrium state.
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
Allele Frequency Analysis

We performed population allele frequency analysis with both InDel multiplexes typing the NIST collection of 712 population samples.

Samples from the four representative human groups of the U.S. population have been used. Unrelated individuals of self-declared ancestry.

- **260 African Americans**
- **262 U.S. Caucasians**
- **140 U.S. Hispanics**
- **50 U.S. Asians**

Working under the assumption of full independence of the markers, the following RMP values were calculated.

<table>
<thead>
<tr>
<th></th>
<th>U.S. Cauc</th>
<th>U.S. Asian</th>
<th>U.S. Hisp</th>
<th>Af-Am</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean DiPplex RMP</td>
<td>1.86E-13</td>
<td>4.67E-11</td>
<td>4.88E-13</td>
<td>5.88E-12</td>
</tr>
<tr>
<td>Mean HID-38plex RMP</td>
<td>3.67E-15</td>
<td>5.11E-14</td>
<td>1.47E-15</td>
<td>4.74E-15</td>
</tr>
</tbody>
</table>
Allele Frequency Analysis

Although both InDel assays mean RMP value is lower than the 13 CODIS STRs, 68 InDel supply discrimination power higher than 20 STRs.
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
Artificially Degraded DNA Assay

We have conducted several experiments in order to produce mimic DNA degradation samples in a controlled way. Only DNA fragmentation processes were simulated.

Our objective is to compare the short-amplicon InDel typing reactions to establish short-amplicon STRs kit performance.

**COVARIS**, A focused acoustic DNA shearing technique now employed in Next Generation Sequencing. The method was applied to create appropriately degraded DNA samples in a controlled fashion. For more information see: http://www.covarisinc.com/how_it_works.htm
Artificially Degraded DNA Assay

Several protocols have been tried before reaching the desired DNA fragmentation (100-250 bp fragments)

Temperature: 5 °C
Mode: Frequency sweeping
Duty Cycle: 10 %
Intensity: 10 %
Cycle/Burst: 1000
Time: 20 minutes
DNA: 50 ng
Dilution volume: 100 μL
Tube: glass- 100 μL tube

Only samples corresponding to these conditions (20') were used for the final analysis
Identifiler – 7 alleles detected

28 PCR cycles

- D8S1179
- D3S1358
- D19S433
- vWA
- Amelogenin
- D5S818
Minifiler – 16 alleles detected

- D13S317
- D7S820
- Amelogenin
- D2S1338
- D21S11
- 230 bps
- D16S539
- D18S51
- CSF1PO
- FGA

Signal progressive decrease on the longer markers

All profiles shown scaled to 2000 RFUs

30 PCR cycles
All profiles shown scaled to 2000 RFUs

DIPplex – 49 alleles detected

Signal progressive decrease on the longer markers

160 bps

30 PCR cycles
HID-38plex – 43 alleles detected

Progressive signal decrease of the longer markers resulted in drop out of 5 of the longest markers.

29 PCR cycles

All profiles shown scaled to 2000 RFUs
Artificially Degraded DNA Assay

With the number of observed alleles on each kit, we obtained the following RMP values:

<table>
<thead>
<tr>
<th>Assay</th>
<th>Exp. Alleles</th>
<th>Obs. Alleles</th>
<th>Loci total</th>
<th>Amp. Loci</th>
<th>RMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifiler</td>
<td>10</td>
<td>5</td>
<td>15</td>
<td>5</td>
<td>n/a</td>
</tr>
<tr>
<td>Minifiler</td>
<td>16</td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>2.89 e^(-12)</td>
</tr>
<tr>
<td>DIPplex</td>
<td>49</td>
<td>49</td>
<td>30</td>
<td>30</td>
<td>4.77 e^(-14)</td>
</tr>
<tr>
<td>HID-38plex</td>
<td>43*</td>
<td>43*</td>
<td>38</td>
<td>33</td>
<td>1.03 e^(-14)</td>
</tr>
</tbody>
</table>

* Based on surviving loci

- We could assure that the application of short amplicon markers such as DIPplex and MiniFiler to challenging DNA samples would be of great interest for real casework.

- In case of limited amount of sample, InDel marker amplification should be considered in spite of other assays, such as Minifiler, unless core STRs are needed.

- For future sample preparation, increased shearing times could be tried in order to achieve a further level of fragmentation.
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
Sequencing of Previously Unreported Variation

Consistent heterozygote peak imbalance

Balanced Heterozygote

Imbalanced Heterozygote

This suggested the presence of a SNP within the primer binding site potentially disrupting primer annealing in samples carrying the minor allele.
• A neighboring SNP (G/A), located 61 bp downstream from the main InDel site. This is a SNP referenced in the dbSNP database as rs17245568. The A allele of this SNP corresponds to the samples carrying the observed imbalance.

• We do not have the Qiagen PCR primer sequences. It is reasonable to assume that the G→A SNP 61 bases downstream from the insertion is the cause of the peak imbalance.
Sequencing of Previously Unreported Variation

-A second feature would be the presence of a seemingly **third off-ladder allele** for two of the DIPplex markers (D99 and D84).

Two possible explanations for these features:

+ A different size deletion/insertion allele at the locus
+ An additional neighboring InDel site with a rare minor allele within the amplicon range.

D84 – rs3081400
off-ladder allele
D84 – rs3081400 off-ladder allele

- Insertion
- Neighboring Insertion
- D84 Insertion allele

- Deletion
- Neighboring Insertion
- D84 Deletion allele

Referenced InDel on dbSNP database as rs11573892.
- We would suggest a reformulation of the reverse primer for the marker D97, as nearly as much as a quarter of the analyzed African-American samples displayed imbalance.

- This situation may lead, especially in degraded DNA samples, to the drop-out of the Insertion allele of this marker.

- The non-standard mobility variants observed in the Qiagen DIPplex InDel set have proven to be stable and due to a single characterized polymorphic variant.

- The characterization of such rarer mobility variants, far from being a hindrance, can further contribute to the informative power of InDel typing.
Presentation Outline

1- InDel Polymorphisms: Introduction and Concept.

2- Materials and Methods:
   - InDel assays HID-38plex and DIPplex.
   - Independence of the markers.

3- Results
   - Allele frequency analysis.
   - Artificially degraded DNA assay.
   - Sequencing of previously unreported variation.

4- Conclusions
Conclusions

- Collected U.S. population data (n=712) on 68 InDel loci (In 2 multiplexes and 1.6 ng of total DNA)

- Demonstrated improved success rates with artificially degraded DNA compared to Identifiler STR typing.

- Characterized some unreported off-ladder alleles and imbalanced heterozygotes.

- InDels can be a supporting tool to STRs for challenging casework samples.
Acknowledgements

- **Qiagen** for providing access to DIPplex kits.
- **Margaret Kline and Becky Hill** from NIST for assistance with allele sequencing.
- **Jennifer McDaniel** from NIST for assistance with the COVARIS system.
- **Carla Santos** from USC for providing HID-38plex primer mix.

A final version of the slides will be uploaded to STRbase webpage http://www.cstl.nist.gov/strbase/ISHI2011-InDel.pdf